

# QUANTIFYING NUISANCE PARAMETER EFFECTS VIA DECOMPOSITIONS OF ASYMPTOTIC REFINEMENTS FOR LIKELIHOOD-BASED STATISTICS

THOMAS J. DICICCIO, TODD A. KUFFNER, AND G. ALASTAIR YOUNG

**ABSTRACT.** Accurate inference on a scalar interest parameter in the presence of a nuisance parameter may be obtained using an adjusted version of the signed root likelihood ratio statistic, in particular Barndorff-Nielsen's  $R^*$  statistic. The adjustment made by this statistic may be decomposed into a sum of two terms, interpreted as correcting respectively for the possible effect of nuisance parameters and the deviation from standard normality of the signed root likelihood ratio statistic itself. We show that the adjustment terms are determined to second-order in the sample size by their means. Explicit expressions are obtained for the leading terms in asymptotic expansions of these means. These are easily calculated, allowing a simple way of quantifying and interpreting the respective effects of the two adjustments, in particular of the effect of a high dimensional nuisance parameter. Illustrations are given for a number of examples, which provide theoretical insight to the effect of nuisance parameters on parametric inference. The analysis provides a decomposition of the mean of the signed root statistic involving two terms: the first has the property of taking the same value whether there are no nuisance parameters or whether there is an orthogonal nuisance parameter, while the second is zero when there are no nuisance parameters. Similar decompositions are discussed for the Bartlett correction factor of the likelihood ratio statistic, and for other asymptotically standard normal pivots.

## 1. INTRODUCTION

We are concerned with inference on a scalar interest parameter in the presence of a, possibly high dimensional, nuisance parameter, based on a data sample of size  $n$ , and with identification of procedures which yield repeated sampling accuracy. In this setting, inference accurate to third order, that is with repeated sampling error of order  $O(n^{-3/2})$ , may be obtained using an adjusted version of the signed root likelihood ratio statistic, in particular through use of Barndorff-Nielsen's  $R^*$  statistic (Barndorff-Nielsen, 1986).

The  $R^*$  statistic is particularly useful in two contexts. In full, multi-parameter exponential family models inference based on standard normal approximation to the sampling distribution of the  $R^*$  statistic approximates to third order the optimal, conditional, but generally intractable, inference, which is based on conditioning on the sufficient statistic for the nuisance parameter. In more general models which admit an ancillary statistic, taken to mean an approximately distribution free statistic which together with the maximum likelihood estimator constitutes a minimal sufficient statistic for the full parameter in the model, the normal approximation approximates to the same third order an exact inference based on conditioning on the ancillary statistic. A practical limitation of the use of  $R^*$  is in the requirement of explicit specification of the appropriate ancillary, and the need to express the likelihood directly in terms of the

---

*Key words and phrases.* Adjusted signed root likelihood ratio; Ancillary Statistic; Bartlett correction; Cornish-Fisher; Decomposition; Exponential family; Nuisance parameter; Profile likelihood.

maximum likelihood estimator and the ancillary statistic. When calculation of the  $R^*$  statistic is tractable, inference with repeated sampling accuracy  $O(n^{-3/2})$  is obtained through the normal approximation. This same level of repeated sampling accuracy may be obtained by parametric bootstrap procedures, in particular those based on simulation estimation of the sampling distribution of the unadjusted signed root statistic: see DiCiccio et al. (2001), Lee & Young (2005). Key to this bootstrap approach is appropriate handling of the nuisance parameter: third order repeated sampling accuracy is obtained by considering the sampling distribution of the signed root statistic when the nuisance parameter is specified as the constrained maximum likelihood value calculated from the observed data sample.

Inference based on the  $R^*$  statistic and the parametric bootstrap alternative sketched above are analytically related. DiCiccio & Young (2008) observe that in the problem of inference on a scalar component of the canonical parameter in the multi-parameter exponential family context, inference based on normal approximation to  $R^*$  may be viewed as an analytic, saddlepoint approximation to the bootstrap inference. In the same way, it is readily seen that in the ancillary statistic context, inference based on  $R^*$  may be regarded as a saddlepoint approximation to a conditional bootstrap calculation, which simulates the distribution of the signed root statistic conditional on the observed value of the ancillary statistic, with the nuisance parameter fixed at its constrained maximum likelihood value. Simulation of this conditional bootstrap distribution will be infeasible in many circumstances, though in certain cases, such as regression-scale models, simple methods of conditional simulation, employing MCMC, are possible: see Brazzale & Davison (2008). Alternatively, and more simply, the conditional distribution may be replaced by simulation of the marginal distribution of the signed root statistic. DiCiccio et al. (2015) demonstrate that the marginal bootstrap distribution approximates the conditional bootstrap distribution to second order,  $O(n^{-1})$ , given the ancillary statistic.

The adjustment made by the  $R^*$  statistic may be decomposed into a sum of two terms, interpreted as correcting respectively for the possible effect of nuisance parameters and an information adjustment, representing the deviation from standard normality of the signed root likelihood ratio statistic itself. Pierce & Peters (1992) proposed such a decomposition in the case where the interest parameter is a component of the canonical parameter in a full exponential family model. A generalization of the decomposition is detailed by Barndorff-Nielsen & Cox (1994, Section 6.6.4). Starting from numerical investigations by Pierce & Peters (1992), it has been noted that the information adjustment is typically small when the adjusted information for the interest parameter, which we define formally in Section 2, is large. By contrast, the nuisance parameter adjustment can be appreciable when information on the nuisance parameter is small, as will usually occur when its dimension is large. Crucially, however, the magnitude of the nuisance parameter adjustment relative to the information adjustment also depends on the structure of the statistical model in question, and a simple methodology for measurement of nuisance parameter effects for a given model is lacking.

In this paper we note that the adjustment terms are, from a repeated sampling perspective, determined to second-order,  $O(n^{-1})$ , in the sample size by their means. The precise definitions of the adjustment terms themselves are unimportant to our strategy for quantifying nuisance parameter effects, though we note that, except for full exponential family and transformation models, they must generally be approximated, leading to only second-order accuracy from the resulting adjusted signed root statistic. Approximations to  $R^*$  which yield second-order accuracy include those described by DiCiccio & Martin (1993) and Skovgaard (1996): for a summary see Severini (2000, Section 7.5).

We obtain explicit expressions for the leading terms in asymptotic expansions of the repeated sampling means of the nuisance parameter and information adjustments. These involve calculation only of expectations of certain low-order log-likelihood derivatives, and are therefore easily evaluated for quite general models, even when the  $R^*$  statistic itself is intractable. The adjustment terms have variances of low order  $O(n^{-2})$  and the asymptotic means therefore allow a simple, effective and general way of quantifying and interpreting the respective effects of the two adjustments. Of particular methodological interest is analysis of the effect of a high dimensional nuisance parameter on the inference based on the  $R^*$  statistic, and by extension its bootstrap alternative. Inference based on the  $R^*$  statistic, when tractable, represents a ‘gold standard’ in what is achievable in the inference problem and we have noted a close relationship between inference based on the  $R^*$  statistic and parametric bootstrap inference. It is reasonable therefore to expect that the calculations are useful too in shedding light on operation of the parametric bootstrap. The repeated sampling properties of the bootstrap are, modulo Monte Carlo error introduced by the need in practice to construct the bootstrap estimate of the sampling distribution of the signed root statistic from a finite simulation, determined entirely by nuisance parameter effects, through substitution of unknown values by estimates. A central recommendation of this paper is that valuable insights to operation of the parametric bootstrap may be obtained by identification of the explicit way in which the means of the nuisance parameter and information adjustments depend on the nuisance parameter. As we shall see in Section 4, in certain key problems these quantities depend only on the dimension of the nuisance parameter, and not on its actual value. In such cases we may reasonably expect good repeated sampling accuracy from the bootstrap, as precise specification of the nuisance parameter values in the calculation is unimportant. In other situations, we observe that the value of the nuisance parameter has a more substantial effect on the adjustment means, in which case we may be alert to impaired accuracy from the bootstrap and its analytic alternatives, especially with small sample sizes.

Our analysis provides a decomposition of the mean of the signed root statistic involving two terms: the first has the property of taking the same value whether there are no nuisance parameters or whether there is an orthogonal nuisance parameter, while the second is zero when there are no nuisance parameters. Similar decompositions are discussed for the Bartlett correction factor of the likelihood ratio statistic, and for other asymptotically standard normal pivots, in Sections 5 and 6 respectively.

## 2. THE INFERENCE PROBLEM

Suppose that  $Y = (Y_1, \dots, Y_n)$  is a continuous random vector and that the distribution of  $Y$  depends on an unknown  $d$ -dimensional parameter  $\theta = (\theta^1, \dots, \theta^d)$ , partitioned as  $\theta = (\psi, \phi)$ , where  $\psi = \theta^1$  is a scalar interest parameter and  $\phi$  is a nuisance parameter of dimension  $d - 1$ . Let  $L(\theta)$  be the log-likelihood function for  $\theta$  based on  $Y$  and let  $\hat{\theta} = (\hat{\psi}, \hat{\phi})$  be the global maximum likelihood estimator of  $\theta$ . Further, let  $\tilde{\theta} = \tilde{\theta}(\psi) = (\psi, \tilde{\phi}) = \{\psi, \tilde{\phi}(\psi)\}$  be the constrained maximum likelihood estimator of  $\theta$  for given  $\psi$ . Then the profile log-likelihood function for  $\psi$  is  $M(\psi) = L\{\tilde{\theta}(\psi)\}$  and the likelihood ratio statistic for  $\psi$  is  $W(\psi) = 2\{M(\hat{\psi}) - M(\psi)\}$ , where  $M(\hat{\psi}) = L(\hat{\theta})$ , since  $\tilde{\theta}(\hat{\psi}) = \hat{\theta}$ . The signed root likelihood ratio statistic is  $R(\psi) = \text{sgn}(\hat{\psi} - \psi)\{W(\psi)\}^{1/2}$ . Then, for example, testing  $H_0 : \psi = \psi_0$  against  $H_a : \psi > \psi_0$  or  $H_a : \psi < \psi_0$  can be based on the test statistic  $R(\psi_0)$ . Asymptotically, as the sample size  $n$  increases, the sampling distribution of  $R(\psi)$  tends to the standard normal distribution.

Specifically,  $R(\psi)$  is distributed as standard normal to first order, to error of order  $O(n^{-1/2})$ . By contrast, the  $R^*$  statistic is distributed as standard normal to error of order  $O(n^{-3/2})$ .

The  $R^*$  statistic is defined by

$$(1) \quad R^*(\psi) = R(\psi) + R(\psi)^{-1} \log(v(\psi)/R(\psi)),$$

where  $v(\psi)$  is given (Barndorff-Nielsen, 1986) by

$$(2) \quad v(\psi) = \left| \frac{L_{;\hat{\theta}}(\hat{\theta}) - L_{;\hat{\theta}}(\tilde{\theta})}{L_{\phi;\hat{\theta}}(\tilde{\theta})} \right| / \{|j_{\phi\phi}(\tilde{\theta})|^{1/2} |j(\hat{\theta})|^{1/2}\}.$$

Here, it is supposed that the log-likelihood function has been written as  $L(\theta; \hat{\theta}, a)$ , with  $(\hat{\theta}, a)$  minimal sufficient and  $a$  ancillary, that is with a distribution which, at least approximately, does not depend on  $\theta$ . Further,

$$L_{;\hat{\theta}}(\theta) \equiv L_{;\hat{\theta}}(\theta; \hat{\theta}, a) = \frac{\partial}{\partial \hat{\theta}} L(\theta; \hat{\theta}, a), \quad L_{\phi;\hat{\theta}}(\theta) \equiv L_{\phi;\hat{\theta}}(\theta; \hat{\theta}, a) = \frac{\partial^2}{\partial \phi \partial \hat{\theta}} L(\theta; \hat{\theta}, a).$$

Also,  $j$  denotes the observed information matrix,  $j(\theta) = (-L_{rs}(\theta))$ , with  $L_{rs}(\theta) = \partial^2 L(\theta) / \partial \theta^r \partial \theta^s$ , and  $j_{\phi\phi}$  denotes its  $(\phi, \phi)$  component. The sampling distribution of  $R^*(\psi)$  is standard normal conditionally on  $a$ , and hence, as noted, unconditionally, to error of third order  $O(n^{-3/2})$ . Note that in a full exponential family model,  $\hat{\theta}$  is already itself sufficient, and no ancillary statistic  $a$  is required. The expression for  $v(\psi)$  given by (2) therefore simplifies somewhat: see, for example, Barndorff-Nielsen & Cox (1994, Example 6.19).

Barndorff-Nielsen & Cox (1994, Section 6.6.4), generalizing Pierce & Peters (1992), introduce quantities  $\text{NP}(\psi)$  and  $\text{INF}(\psi)$ , both of order  $O_p(n^{-1/2})$ , such that  $R^*(\psi) = R(\psi) + \text{NP}(\psi) + \text{INF}(\psi)$ . Explicitly, we have

$$\text{NP}(\psi) = -\frac{1}{R(\psi)} \log C(\psi),$$

where

$$C(\psi) = \frac{\{|j_{\phi\phi}(\hat{\theta})| |j_{\phi\phi}(\tilde{\theta})|\}^{1/2}}{|L_{\phi;\hat{\theta}}(\tilde{\theta})|},$$

with  $L_{\phi;\hat{\theta}}(\theta) \equiv L_{\phi;\hat{\theta}}(\theta; \hat{\theta}, a) = \partial^2 L(\theta; \hat{\theta}, a) / \partial \phi \partial \hat{\theta}$  and, as before,  $j_{\phi\phi}$  denoting the  $(\phi, \phi)$  component of the observed information  $j$ . Also,

$$\text{INF}(\psi) = \frac{1}{R(\psi)} \log \{u(\psi)/R(\psi)\},$$

where

$$u(\psi) = j_p(\hat{\psi})^{-1/2} \frac{\partial}{\partial \hat{\psi}} \{M(\hat{\psi}) - M(\psi)\}.$$

Here  $j_p$  is the profile observed information,  $j_p(\psi) = -\partial^2 M(\psi) / \partial \psi^2$ , and the derivative with respect to  $\hat{\psi}$  is calculated with  $M(\hat{\psi}) - M(\psi)$  considered as a function of  $\psi, \hat{\psi}, \tilde{\phi}(\psi)$  and  $a$ .

Calculation of  $R^*(\psi)$  supposes explicit representation of the log-likelihood as a function of  $(\hat{\theta}, a)$ . Other formulations of the adjustment  $v(\psi)$ , due to Fraser and co-workers, are possible. The tangent exponential model introduced by Fraser (1990) avoids the need to specify the transformation  $Y \rightarrow (\hat{\theta}, a)$ , though still requires awkward analytic calculation: a useful summary is given by Brazzale et al. (2007, Chapter 8). In general, however, it is necessary to approximate to the quantity  $v(\psi)$ . Replacing  $v(\psi)$  in the definition (1) of  $R^*(\psi)$  by an estimate  $\tilde{v}(\psi)$  typically yields an adjusted version of the signed root likelihood ratio statistic distributed as standard normal only to error of second order,  $O(n^{-1})$ . A computationally attractive approximation based on orthogonal parameterisation (Cox & Reid, 1987) is described by DiCiccio & Martin (1993). The approximation due to Skovgaard (1996) is theoretically attractive in that it also provides large deviations protection.

To develop our analysis, some further notation is required. Let  $L_\theta(\theta)$  denote the score function, the vector with components  $L_r(\theta) = \partial L(\theta)/\partial \theta^r$ ,  $r = 1, \dots, d$ . In the calculations that follow, arrays and summation are denoted by using the standard conventions, for which the indices  $r, s, t, \dots$  are assumed to range over  $1, \dots, d$ . Summation over the range is implied for any index appearing in an expression both as a subscript and as a superscript. As above, differentiation is indicated by subscripts. Then  $E\{L_r(\theta)\} = 0$ ; let  $\lambda_{rs} = E\{L_{rs}(\theta)\}$ ,  $\lambda_{rst} = E\{L_{rst}(\theta)\}$ , etc., and put  $l_r = L_r(\theta)$ ,  $l_{rs} = L_{rs}(\theta) - \lambda_{rs}$ ,  $l_{rst} = L_{rst}(\theta) - \lambda_{rst}$ , etc. The constants  $\lambda_{rs}$ ,  $\lambda_{rst}$ ,  $\dots$ , are assumed to be of order  $O(n)$ . The variables  $l_r$ ,  $l_{rs}$ ,  $l_{rst}$ , etc., each of which have expectation 0, are assumed to be of order  $O_p(n^{1/2})$ . The joint cumulants of  $l_r$ ,  $l_{rs}$ , etc. are assumed to be of order  $O(n)$ . These assumptions will usually be satisfied in situations involving independent observations, or structured dependence, such as in time series contexts. It is useful to extend the  $\lambda$ -notation: let  $\lambda_{r,s} = E(L_r L_s) = E(l_r l_s)$ ,  $\lambda_{rs,t} = E(L_{rs} L_t) = E(l_{rs} l_t)$ , etc. Bartlett identities involving the  $\lambda$ 's can be derived by repeated differentiation of the identity  $\int \exp\{L(\theta)\} dy = 1$ ; in particular,

$$\lambda_{rs} + \lambda_{r,s} = 0, \quad \lambda_{rst} + \lambda_{rs,t} + \lambda_{rt,s} + \lambda_{st,r} + \lambda_{r,s,t} = 0.$$

Differentiation of the definition  $\lambda_{rs} = \int L_{rs}(\theta) \exp\{L(\theta)\} dy$  yields  $\lambda_{rs/t} = \lambda_{rst} + \lambda_{rs,t}$ , where  $\lambda_{rs/t} = \partial \lambda_{rs} / \partial \theta^t$ . Further, let  $(\lambda^{rs})$  be the  $d \times d$  matrix inverse of  $(\lambda_{rs})$ , and let  $\eta = -1/\lambda^{11}$ ,  $\tau^{rs} = \eta \lambda^{1r} \lambda^{1s}$ , and  $\nu^{rs} = \lambda^{rs} + \tau^{rs}$ . Thus,  $\lambda^{rs}$ ,  $\tau^{rs}$ , and  $\nu^{rs}$  are of order  $O(n^{-1})$ , while  $\eta$ , which is what we have termed the adjusted information for  $\psi$ , is of order  $O(n)$ .

DiCiccio & Stern (1994a) showed that  $R(\psi) = \eta^{1/2}\{R_1 + R_2 + O_p(n^{-3/2})\}$ , where  $R_1 = -\lambda^{1r} l_r$  and

$$R_2 = \lambda^{1r} \lambda^{st} l_{rs} l_t + \frac{1}{2} \lambda^{1r} \tau^{st} l_{rs} l_t - \frac{1}{2} \lambda^{1r} \lambda^{su} \nu^{tv} \lambda_{rst} l_u l_v - \frac{1}{6} \lambda^{1r} \tau^{su} \tau^{tv} \lambda_{rst} l_u l_v.$$

Note that  $R_1$  is of order  $O_p(n^{-1/2})$  and  $R_2$  is of order  $O_p(n^{-1})$ . Since  $E(R_1) = 0$ , it follows that

$$(3) \quad E\{R(\psi)\} = \eta^{1/2}\{\lambda^{1r} \lambda^{st} \lambda_{rs,t} + \frac{1}{2} \lambda^{1r} \tau^{st} \lambda_{rs,t} + \frac{1}{2} \lambda^{1r} \lambda^{st} \lambda_{rst} + \frac{1}{3} \lambda^{1r} \tau^{st} \lambda_{rst}\} + O(n^{-1}).$$

### 3. EXPECTATIONS OF ADJUSTMENTS

Detailed analysis given in the Appendix shows that we may approximate  $E\{\text{INF}(\psi)\}$  to  $O(n^{-1})$  by

$$g_{\text{INF}}(\theta) = \eta^{1/2} \lambda^{1r} \tau^{st} \left( \frac{1}{2} \lambda_{rs,t} + \frac{1}{6} \lambda_{rst} \right),$$

and  $E\{\text{NP}(\psi)\}$  to the same order by

$$g_{\text{NP}}(\theta) = -\eta^{1/2} \lambda^{1r} \nu^{st} (\lambda_{rs,t} + \frac{1}{2} \lambda_{rst}).$$

These expansions permit a full statistical interpretation of the adjustment terms  $\text{NP}(\psi)$  and  $\text{INF}(\psi)$ , which we do through a series of remarks.

*Remark 1.* We begin by examining  $E\{R(\psi)\}$  when there are no nuisance parameters. If nuisance parameters are absent, then  $\lambda^{11} = (\lambda_{11})^{-1}$ ,  $\eta = -\lambda_{11}$ ,  $\tau^{11} = (-\lambda_{11})^{-1}$ , and  $\nu^{11} = 0$ , and it follows that

$$E\{R(\psi)\} = (-\lambda_{11})^{-3/2}(\frac{1}{2}\lambda_{11,1} + \frac{1}{6}\lambda_{111}) + O(n^{-1}).$$

*Remark 2.* The quantities  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  are related to asymptotic quantities detailed by Efron (1987) in description of the ‘bias corrected accelerated’,  $BC_a$ , method of construction of bootstrap confidence intervals, which is analysed in detail by DiCiccio & Efron (1996). Specifically, we have  $g_{\text{INF}}(\theta) = a_0$  and  $g_{\text{NP}}(\theta) = z_0 - a_0$ , where  $a_0 = a_0(\theta)$  and  $z_0 = z_0(\theta)$  are respectively acceleration and bias-correction quantities. The quantity  $a_0$  satisfies (DiCiccio & Efron, 1996)

$$a_0 = -\frac{1}{6}\{\text{skew}(U) + \text{skew}(T)\} + O(n^{-1}),$$

where  $U = (\hat{\psi} - \psi)/\sigma$ , with  $\sigma^2$  the variance of  $\hat{\psi}$ , given by  $\sigma^2 \equiv \sigma^2(\theta) = \lambda^{1,1} + O(n^{-2})$ , and  $T = (\hat{\psi} - \psi)/\hat{\sigma}$ , with  $\hat{\sigma}^2 = \sigma^2(\hat{\theta})$ . Further,  $z_0$  is interpreted by

$$\Phi(z_0) = \Pr(\hat{\psi} \leq \psi) + O(n^{-1}),$$

where  $\Phi$  is the standard normal distribution function.

DiCiccio & Efron (1996) note that the quantities  $a_0$  and  $z_0$  are invariant under reparameterisations of the model. Therefore, in using the asymptotic adjustment expectations  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  to interpret nuisance parameter effects on the inference on  $\psi$ , there is no restriction in assuming that the model under analysis is parameterised so that the interest parameter  $\psi$  and the nuisance parameter  $\phi$  are orthogonal (Cox & Reid, 1987). Therefore, now suppose there is a vector nuisance parameter  $\phi$  present, but assume that the interest parameter  $\psi$  and the nuisance parameter  $\phi$  are orthogonal; then  $\lambda^{11} = (\lambda_{11})^{-1}$ ,  $\eta = -\lambda_{11}$ ,  $\lambda^{1a} = 0$  ( $a = 2, \dots, d$ ),  $\tau^{rs} = 0$  except when  $r = s = 1$ , in which case  $\tau^{11} = (-\lambda_{11})^{-1}$ , and

$$E\{\text{INF}(\psi)\} = -(-\lambda_{11})^{-3/2}(\frac{1}{2}\lambda_{11,1} + \frac{1}{6}\lambda_{111}) + O(n^{-1}).$$

Therefore, following Remark 1, to error of order  $O(n^{-1})$ ,  $E\{\text{INF}(\psi)\}$  is seen to correspond to a mean adjustment for the signed root statistic  $R(\psi)$  in the problem where the orthogonal nuisance parameter  $\phi$  is known. Since the standard normal approximation to the distribution of  $R(\psi)$  is typically rather accurate in scalar parameter cases without nuisance parameters, the mean adjustment should be quantitatively small quite generally, so we can anticipate that  $\text{INF}(\psi)$  is typically small.

*Remark 3.* For general parameterisations, we have  $\nu^{11} = \nu^{a1} = \nu^{1b} = 0$  for  $a, b = 2, \dots, d$ , and thus,

$$\begin{aligned} E\{\text{NP}(\psi)\} &= -\eta^{1/2}\lambda^{1r}\nu^{ab}(\lambda_{ra,b} + \frac{1}{2}\lambda_{rab}) + O(n^{-1}) \\ &= \eta^{1/2}\lambda^{1r}\nu^{ab}(\frac{1}{2}\lambda_{rab} - \lambda_{ra/b}) + O(n^{-1}), \end{aligned}$$

where  $\lambda_{ra/b} = \partial\lambda_{ra}/\partial\theta^b$  and  $\lambda_{ra,b} = \lambda_{ra,b} + \lambda_{rab}$ .

Under orthogonality,  $\nu^{ab} = \lambda^{ab}$  for  $a, b = 2, \dots, d$ , and the condition  $\lambda_{1a} = 0$  for  $a = 2, \dots, d$  implies that  $\lambda_{1a/b} = 0$  for  $b = 2, \dots, d$ , so that the identity  $\lambda_{1a/b} = \lambda_{1a,b} + \lambda_{1ab}$  yields  $\lambda_{1a,b} = -\lambda_{1ab}$  for



$a, b = 2, \dots, d$ . Hence, nuisance parameter effects may be quantified from the expression

$$E\{\text{NP}(\psi)\} = -\frac{1}{2}(-\lambda_{11})^{-1/2}\lambda^{ab}\lambda_{ab1} + O(n^{-1}).$$

Note that this gives  $\beta_1 = \eta^{1/2}E\{\text{NP}(\psi)\} + O(n^{-1/2}) = -\frac{1}{2}\lambda^{ab}\lambda_{ab1} + O(n^{-1/2})$ . Since the expansion for  $E\{\text{NP}(\psi)\}$  involves a multiple sum over the nuisance parameters, we see that  $\text{NP}(\psi)$  can be anticipated to be large when the number of nuisance parameters is large.

*Remark 4.* Some further insight into  $\text{NP}(\psi)$  in the orthogonal case can be gleaned by noting that

$$\frac{\partial \log \det[-L_{ab}\{\tilde{\theta}(\psi)\}]}{\partial \psi} = L^{ab}(\theta)L_{ab1}(\theta) + O_p(n^{-1/2}) = \lambda^{ab}\lambda_{ab1} + O_p(n^{-1/2}),$$

which further relates  $E\{\text{NP}(\psi)\}$  to the specific adjustment function of Cox & Reid (1987). Thus, in this orthogonal case, if  $\log \det\{-L_{ab}(\theta)\}$  does not change rapidly with  $\psi$ , such as when  $L(\theta) = g(\psi) + h(\phi)$ , in which case  $\det\{-L_{ab}(\theta)\}$  is constant with respect to  $\psi$ , then  $\lambda^{ab}\lambda_{ab1}$  is small in magnitude, and hence, we would expect  $\text{NP}(\psi)$  to be small in magnitude; see also the discussion in Cox & Reid (1987).

*Remark 5.* There is one further interpretation of  $\text{NP}(\psi)$  that is worth noting. DiCiccio & Stern (1994a) showed that the difference between  $\bar{\psi}$  and  $\hat{\psi}$  is

$$\bar{\psi} - \hat{\psi} = -\lambda^{11}\beta_1 + O_p(n^{-3/2}) = \eta^{-1}\beta_1 + O_p(n^{-3/2}) = \eta^{-1/2}E\{\text{NP}(\psi)\} + O_p(n^{-3/2}),$$

and hence, this difference, when expressed in terms of standard deviations of  $\hat{\psi}$ , is

$$\frac{\bar{\psi} - \hat{\psi}}{\eta^{-1/2}} = E\{\text{NP}(\psi)\} + O_p(n^{-1}).$$

*Remark 6.* Note that the quantities  $g_{\text{NP}}(\theta)$  and  $g_{\text{INF}}(\theta)$  are both of order  $O(n^{-1/2})$ . As we shall illustrate, calculation of the individual values provides important statistical insight. We propose further that a simple measure of the relative influence within the assumed model of the nuisance parameter on inference on the interest parameter  $\psi$ , independent of the sample size  $n$ , might be obtained by considering their ratio  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta)$ .

*Remark 7.* In general, the quantities  $g_{\text{NP}}(\theta)$  and  $g_{\text{INF}}(\theta)$  depend on the unknown parameter  $\theta$ . In practice, following the bootstrap principle, they may be estimated by  $g_{\text{NP}}(\tilde{\theta})$  and  $g_{\text{INF}}(\tilde{\theta})$  respectively. An adjusted version of the signed root statistic  $R(\psi)$ , easily calculated in practice, once  $g_{\text{NP}}(\theta)$  and  $g_{\text{INF}}(\theta)$  have been calculated, is given by  $R_a(\psi) = R(\psi) + g_{\text{NP}}(\tilde{\theta}) + g_{\text{INF}}(\tilde{\theta})$ . Since  $g_{\text{NP}}(\tilde{\theta}) - g_{\text{NP}}(\theta) = O_p(n^{-1})$ , we have that  $R_a(\psi) = R^*(\psi) + O_p(n^{-1})$ , and therefore that  $R_a(\psi)$  has the standard normal distribution to error of order  $O(n^{-1})$ . DiCiccio & Efron (1996) previously remarked that  $R(\psi) + z_0(\hat{\theta})$  is standard normal to error of order  $O(n^{-1})$ , but did not investigate practical use of this statistic for inference: an alternative is the statistic  $R_a(\psi) = R(\psi) + z_0(\tilde{\theta})$ . Although no claim of desirable large deviation properties of the kind enjoyed by the method of Skovgaard (1986) can be made for this statistic, empirical evidence, not reported here, suggests that it nevertheless yields highly accurate inference in many settings.

*Remark 8.* Note that the asymptotic regime adopted here is one in which the dimensionality  $d - 1$  of the nuisance parameter  $\phi$  remains fixed as the sample size  $n$  increases. However, we propose that examination of the quantities  $g_{\text{NP}}(\theta)$  and  $g_{\text{INF}}(\theta)$  and their ratio is a useful device to quantify the effect of an increasing dimension of nuisance parameter on the inference, as we shall illustrate in the next Section. For stratified models, such as those in Examples 2, 4, 5 and 6 below, Sartori (2003) noted

that, when both the sample size  $n$  within each stratum and the number of nuisance parameters  $q$  tend to infinity,  $NP(\psi) = O_p(qm^{-1/2})$ , while  $INF(\psi) = O_p(m^{-1/2})$ , where  $m = nq$  is the total sample size, irrespective of the nature of the sequence  $\{q, n\}$ . Hence, the ratio  $NP(\psi)/INF(\psi) = O_p(q)$  in such an asymptotic regime, consistent with calculations given in Examples 2, 4, 5 and 6 below. Relative to the inference adjustment, the nuisance parameter adjustment increases at a rate proportional to the dimension of the nuisance parameter.

#### 4. EXAMPLES

We consider here a number of theoretical and numerical examples.

**Example 1. Normal linear regression.** Let  $Y_1, \dots, Y_n$  denote independent random variables of the form  $Y_i = x_i^T \beta + \sigma \epsilon_i$ , where  $x_1, \dots, x_n$  are known covariate vectors of length  $q$ ,  $\sigma$  is an unknown scalar interest parameter and  $\beta$  is an unknown nuisance parameter vector of length  $q$ , so that  $\theta = (\sigma, \beta)$ . The  $\epsilon_i$  are assumed to be independent standard normal random variables.

In this case,  $n^{1/2}g_{INF}(\theta) = 2^{1/2}/3$  and  $n^{1/2}g_{NP}(\theta) = q/2^{1/2}$ . Note that these quantities do not depend on the parameter value  $\theta$ , while  $\eta = 2n/\sigma^2$ . Nuisance parameter effects are determined, to second order, only by the dimensionality of the nuisance parameter  $\beta$ , not its value. This observation in turn would suggest that inference based on the bootstrap distribution of  $R(\sigma)$  should be highly accurate. In fact,  $R(\sigma)$  is a simple function of  $\hat{\sigma}^2/\sigma^2$ , which has a distribution free of  $\theta$ :  $(n - q)\hat{\sigma}^2/\sigma^2$  is distributed as chi-squared on  $n - q$  degrees of freedom. A bootstrap calculation will, modulo simulation variability, reproduce the exact sampling distribution of  $R(\sigma)$ .

**Example 2. Neyman-Scott model.** Let  $Y_{ij}$ , for  $i = 1, \dots, n$  and  $j = 1, \dots, q$  be independent Gaussian random variables, with  $Y_{ij}$  being distributed as  $N(\mu_j, \sigma^2)$ . The interest parameter is  $\sigma$ , with nuisance parameter  $(\mu_1, \dots, \mu_q)$ , so that  $\theta = (\sigma, \mu_1, \dots, \mu_q)$ .

Now we calculate  $n^{1/2}g_{INF}(\theta) = 1/\{1.5(2q)^{1/2}\}$ , with  $n^{1/2}g_{NP}(\theta) = (q/2)^{1/2}$ , so that  $g_{NP}(\theta)/g_{INF}(\theta) = 1.5q$ . Again, these quantities do not depend on the value of  $\theta$ , only the dimension  $q$  of the nuisance parameter. The adjusted information is given by  $\eta = 2nq/\sigma^2$ . As in Example 1, the signed root statistic  $R(\sigma)$  has a distribution free of the parameter value: it is a function of the pivotal quantity  $\hat{\sigma}^2/\sigma^2$ , and its exact sampling distribution can be constructed by bootstrapping.

A related problem concerns a generalisation of the Behrens-Fisher problem, in which we observe  $Y_{ij}$ , for  $i = 1, \dots, n$  and  $j = 1, \dots, q$  to be independent Gaussian random variables, with  $Y_{ij}$  being distributed as  $N(\mu, \sigma_j^2)$ . The interest parameter is the common mean  $\mu$ , with  $(\sigma_1^2, \dots, \sigma_q^2)$  as nuisance. In this case, we see that  $E\{INF(\psi)\}$  and  $E\{NP(\psi)\}$  are both  $O(n^{-1})$ , not  $O(n^{-1/2})$ . Nuisance parameter effects are quantitatively slight though, by contrast with what is noted above, in this case the signed root statistic  $R(\mu)$  is not exactly pivotal, and the bootstrap inference is not exact. Limited numerical results given by Young (2009) for the case  $q = 2$  would indicate, however, that the bootstrap inference is highly accurate even for small sample size  $n$ .

**Example 3. Exponential regression.** Suppose  $Y_1, \dots, Y_n$  are independent exponential random variables, with means depending on given covariate values. We suppose for simplicity the case of two covariates, though our conclusions extend immediately to the case with a general number of covariates. So, we suppose  $Y_i$  is exponentially distributed with mean  $\phi_1 \exp(-\psi z_i - \phi_2 w_i)$ , with  $\sum z_i = \sum w_i = 0$ , and  $\psi$  the interest parameter. Routine calculations show that  $g_{INF}(\theta)$  and  $g_{NP}(\theta)$ , though complicated



functions of the covariate values, are again free of the parameter  $\theta = (\psi, \phi_1, \phi_2)$ . Further, the signed root statistic  $R(\psi)$  is again easily seen to be exactly pivotal, and bootstrap inference is once more exact.

In the simple case of a single covariate, with  $E(Y_i) = \phi \exp(-\psi z_i)$ , with  $\sum z_i = 0$ , we have

$$E\{\text{NP}(\psi)\} = 0 + O(n^{-1}), \quad E\{\text{INF}(\psi)\} = -(\sum z_i^2)^{-3/2}(\frac{1}{6} \sum z_i^3) + O(n^{-1}) :$$

the nuisance parameter adjustment has expectation of smaller order of magnitude than that of the information adjustment.

We consider now from a numerical perspective three examples with many nuisance parameters previously discussed by Sartori et al. (1999). In each, we provide illustration of dependence of the measure  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta)$  on the dimensionality of the nuisance parameter.

**Example 4. Inverse Gaussian model.** Let  $Y_{ij}$ , for  $i = 1, \dots, n$  and  $j = 1, \dots, q$  be independent, inverse Gaussian random variables, with  $Y_{ij}$  having probability density

$$f(y; \psi, \phi_j) = \{\psi/(2\pi)\}^{1/2} y^{-3/2} \exp\{-\frac{1}{2}(\psi y^{-1} + \phi_j y) + (\psi \phi_j)^{1/2}\}, \quad y > 0,$$

where  $\psi > 0$  and  $\phi_j > 0$ , so that  $\theta = (\psi, \phi_1, \dots, \phi_q)$  and the overall sample size is  $m = nq$ .

Simple algebraic manipulations show that, independently of the parameter value  $\theta$ ,  $n^{1/2}g_{\text{INF}}(\theta) = -1/\{1.5(2q)^{1/2}\}$ , and  $n^{1/2}g_{\text{NP}}(\theta) = -(q/2)^{1/2}$ , so that  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta) = 1.5q$  in this model. We note that in this model the adjusted information for  $\psi$  is given by  $\eta = nq/(2\psi^2)$ .

**Example 5. Multi-sample exponential model.** Let  $Y_{ij}$ , for  $i = 1, \dots, n$  and  $j = 1, \dots, q$  be independent, exponential random variables, with  $Y_{ij}$  having mean  $1/\phi_j$ . The parameter of interest is

$$\psi = q^{-1} \sum_{j=1}^q \exp(-\phi_j t_0),$$

where  $t_0 > 0$  is a fixed constant and  $\theta = (\psi, \phi)$ , with the nuisance parameter  $\phi = (\phi_2, \dots, \phi_q)$ . As noted by Sartori et al. (1999),  $q\psi$  may be interpreted as the expected number of items failing by  $t_0$  in a parallel system with failures rates  $\phi_1, \dots, \phi_q$ .

The interest parameter  $\psi$  is therefore a nonlinear function of the canonical parameter in a full exponential family model. Again, construction of the information and nuisance parameter adjustments  $\text{INF}(\psi)$  and  $\text{NP}(\psi)$  is straightforward, though the constrained maximum likelihood estimator  $\tilde{\theta}$  must be calculated numerically.

By contrast with previous examples, in this model the ratio  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta)$  depends on the value of the parameter  $\theta$ . Values illustrating the effect of increasing nuisance parameter dimension are given in Table 1 for two cases. In both  $t_0=0.5$ : case (a) considers  $\phi_i = 1, i = 1, \dots, q$ , so that  $\psi = 0.6065$ ; case (b) fixes  $\psi = 0.0333$  for each dimension of nuisance parameter, sets  $\exp(-\phi_q t_0) = q\psi/2$  and fixes  $\phi_1 = \dots = \phi_{q-1}$ , the common value being determined by the specified  $\psi$ . Acute dependence of the ratio on the actual parameter values, rather than just the nuisance parameter dimension as in previous examples, is apparent.

**Example 6. Curved exponential family model.** Our final example concerns a model for which calculation of  $R^*(\psi)$  is intractable: the sample space derivatives, derivatives of the log-likelihood with respect to the maximum likelihood estimator, required by the construction (2) of  $R^*(\psi)$ , must be approximated.

TABLE 1. Dependence of ratio  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta)$  on  $q$ , multi-sample exponential model. Case (a) has  $\phi_i = 1, i = 1, \dots, q$ , case (b) has  $\phi_1 = \dots = \phi_{q-1}$ , with  $\exp(-\phi_q t_0) = q\psi/2$ .

$q$	2	5	10	20	50
(a)	2.25	9.00	20.25	42.75	110.25
(b)	-2.10	-5.50	-8.56	-15.76	-130.29

TABLE 2. Dependence of ratio  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta)$  on  $q$ , multi-sample curved exponential family model. Case (a) has  $\psi = 1, \mu_i = i, i = 1, \dots, q$ , case (b) has  $\psi = 1, \mu_i = 1, i = 1, \dots, q$ .

$q$	1	2	5	10	20	50
(a)	1.11	2.45	6.77	14.17	29.09	74.01
(b)	1.11	2.21	5.53	11.05	22.11	55.26

By contrast, the calculations required to evaluate  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  are no more complex than in the other examples.

Let  $Y_{ij}$ , for  $i = 1, \dots, n$  and  $j = 1, \dots, q$  be independent normal random variables with means  $\mu_j > 0$  and variances  $\psi\mu_j^{1/2}$ . This model constitutes a curved exponential family. The parameter of interest is  $\psi$ , with  $\mu_1, \dots, \mu_q$  as nuisance parameters,  $\theta = (\psi, \mu_1, \dots, \mu_q)$ .

Again, the ratio  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta)$  depends on the value of the parameter  $\theta$ . Illustrative values are given in Table 2, for two cases: case (a) has  $\psi = 1, \mu_i = i, i = 1, \dots, q$ , while case (b) has  $\psi = 1, \mu_i = 1, i = 1, \dots, q$ .

## 5. DECOMPOSITION OF THE BARTLETT CORRECTION FACTOR

Recall that the sum of  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  is, to  $O(n^{-1})$ , equal to

$$\begin{aligned} E\{-R(\psi)\} &= -\eta^{1/2}(\lambda^{1r}\lambda^{st}\lambda_{rs,t} + \frac{1}{2}\lambda^{1r}\tau^{st}\lambda_{rs,t} + \frac{1}{2}\lambda^{1r}\lambda^{st}\lambda_{rst} + \frac{1}{3}\lambda^{1r}\tau^{st}\lambda_{rst}) \\ &= -\eta^{1/2}\lambda^{1r}\lambda^{st}(\lambda_{rs,t} + \frac{1}{2}\lambda_{rst}) - \eta^{1/2}\lambda^{1r}\tau^{st}(\frac{1}{2}\lambda_{rs,t} + \frac{1}{3}\lambda_{rst}). \end{aligned}$$

To decide how we might choose  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  in a decomposition of this sum, consider imposing two conditions: first,  $g_{\text{INF}}(\theta)$  must take the same value whether we have no nuisance parameters or we have orthogonal nuisance parameters; and second,  $g_{\text{NP}}(\theta)$  must be 0 when we have no nuisance parameters. These conditions suggest that  $\tau^{rs}$  and  $\nu^{rs}$  play a key role. Note that  $\tau^{11} = (-\lambda_{11})^{-1}$  when there are no nuisance parameters, while for orthogonal nuisance parameters  $\tau^{rs} = 0$  except when  $r = s = 1$ , in which case  $\tau^{11} = (-\lambda_{11})^{-1}$ . Thus,  $\tau^{rs}$  is the same in the orthogonal nuisance parameter case as it is when nuisance parameters are absent. On the other hand, since  $\nu^{rs} = 0$  whenever either or both of  $r$  and  $s$  are 1, we have that  $\nu^{11} = 0$  when there are no nuisance parameters. It is readily seen that the decomposition of the sum into  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  according to the two conditions can be achieved if we substitute  $\lambda^{st} = \nu^{st} - \tau^{st}$  in the sum and then take  $g_{\text{INF}}(\theta)$  to consist of those terms involving  $\tau^{st}$  and

take  $g_{\text{NP}}(\theta)$  to consist of those terms involving  $\nu^{st}$ . We demonstrate here that the same reasoning may be applied to obtain a decomposition of the Bartlett correction factor for the likelihood ratio statistic  $W(\psi)$ .

Lawley (1956) showed (see also DiCiccio & Stern, 1994a) that the expectation of  $W(\psi)$  is  $E\{W(\psi)\} = 1 + b(\theta) + O(n^{-3/2})$ , where

$$\begin{aligned} b(\theta) = & (\lambda^{rs}\lambda^{tu} - \nu^{rs}\nu^{tu})\left(\frac{1}{4}\lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su}\right) \\ & - (\lambda^{rs}\lambda^{tu}\lambda^{vw} - \nu^{rs}\nu^{tu}\nu^{vw})\left(\frac{1}{4}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}\right) \\ & - (\lambda^{ru}\lambda^{sw}\lambda^{tv} - \nu^{ru}\nu^{sw}\nu^{tv})\left(\frac{1}{6}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}\right). \end{aligned}$$

We now decompose  $b(\theta)$  into the sum  $b(\theta) = b_{\text{INF}}(\theta) + b_{\text{NP}}(\theta)$ , where  $b_{\text{INF}}(\theta)$  is the same whether we have no nuisance parameters or whether we have orthogonal nuisance parameters, and  $b_{\text{NP}}(\theta)$  is 0 when there are no nuisance parameters. We make the substitution  $\lambda^{rs} = \nu^{rs} - \tau^{rs}$  in  $b(\theta)$ :  $b_{\text{INF}}(\theta)$  consists of those terms involving the  $\tau^{rs}$  but not the  $\nu^{rs}$ ;  $b_{\text{NP}}(\theta)$  consists of those terms that involve the  $\nu^{rs}$  in any way.

Succinct expressions for  $b_{\text{INF}}(\theta)$  and  $b_{\text{NP}}(\theta)$  derived this way are

$$\begin{aligned} b_{\text{INF}}(\theta) = & \tau^{rs}\tau^{tu}\left(\frac{1}{4}\lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su}\right) \\ & + \tau^{rs}\tau^{tu}\tau^{vw}\left(\frac{1}{4}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}\right) \\ & + \tau^{ru}\tau^{sw}\tau^{tv}\left(\frac{1}{6}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}\right), \end{aligned}$$

and

$$\begin{aligned} b_{\text{NP}}(\theta) = & (\lambda^{rs}\lambda^{tu} - \tau^{rs}\tau^{tu} - \nu^{rs}\nu^{tu})\left(\frac{1}{4}\lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su}\right) \\ & - (\lambda^{rs}\lambda^{tu}\lambda^{vw} + \tau^{rs}\tau^{tu}\tau^{vw} - \nu^{rs}\nu^{tu}\nu^{vw})\left(\frac{1}{4}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}\right) \\ & - (\lambda^{ru}\lambda^{sw}\lambda^{tv} + \tau^{ru}\tau^{sw}\tau^{tv} - \nu^{ru}\nu^{sw}\nu^{tv})\left(\frac{1}{6}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}\right). \end{aligned}$$

If there are no nuisance parameters or there are orthogonal nuisance parameters, then

$$\begin{aligned} b_{\text{INF}}(\theta) = & (\lambda_{11})^{-2}\left(\frac{1}{4}\lambda_{1111} - \lambda_{111/1} + \lambda_{11/11}\right) \\ & - (\lambda_{11})^{-3}\left(\frac{1}{4}\lambda_{111}\lambda_{111} - \lambda_{111}\lambda_{11/1} + \lambda_{11/1}\lambda_{11/1}\right) \\ & - (\lambda_{11})^{-3}\left(\frac{1}{6}\lambda_{111}\lambda_{111} - \lambda_{111}\lambda_{11/1} + \lambda_{11/1}\lambda_{11/1}\right). \end{aligned}$$

Note that if there are no nuisance parameters,  $\nu^{11} = 0$  and  $\tau^{11} = -\lambda^{11}$ , so that  $b_{\text{NP}}(\theta)$  is identically zero. It is useful to evaluate  $b_{\text{NP}}(\theta)$  in the case of orthogonal nuisance parameters to show better the effect of nuisance parameters. Now, by making the substitution  $\lambda^{rs} = \nu^{rs} - \tau^{rs}$ , we have

$$\begin{aligned} b_{\text{NP}}(\theta) = & \{(\nu^{rs} - \tau^{rs})(\nu^{tu} - \tau^{tu}) - \tau^{rs}\tau^{tu} - \nu^{rs}\nu^{tu}\}\left(\frac{1}{4}\lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su}\right) \\ & - \{(\nu^{rs} - \tau^{rs})(\nu^{tu} - \tau^{tu})(\nu^{vw} - \tau^{vw}) + \tau^{rs}\tau^{tu}\tau^{vw} - \nu^{rs}\nu^{tu}\nu^{vw}\} \\ & \quad \times \left(\frac{1}{4}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}\right) \\ & - \{(\nu^{ru} - \tau^{ru})(\nu^{sw} - \tau^{sw})(\nu^{tv} - \tau^{tv}) + \tau^{ru}\tau^{sw}\tau^{tv} - \nu^{ru}\nu^{sw}\nu^{tv}\} \\ & \quad \times \left(\frac{1}{6}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}\right) \end{aligned}$$

$$\begin{aligned}
&= -(\tau^{rs}\nu^{tu} + \nu^{rs}\tau^{tu})(\frac{1}{4}\lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su}) \\
&\quad - (\tau^{rs}\tau^{tu}\nu^{vw} + \tau^{rs}\nu^{tu}\tau^{vw} + \nu^{rs}\tau^{tu}\tau^{vw} - \tau^{rs}\nu^{tu}\nu^{vw} - \nu^{rs}\tau^{tu}\nu^{vw} - \nu^{rs}\nu^{tu}\tau^{vw}) \\
&\quad \times (\frac{1}{4}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}) \\
&\quad - (\tau^{ru}\tau^{sw}\nu^{tv} + \tau^{ru}\nu^{sw}\tau^{tv} + \nu^{ru}\tau^{sw}\tau^{tv} - \tau^{ru}\nu^{sw}\nu^{tv} - \nu^{ru}\tau^{sw}\nu^{tv} - \nu^{ru}\nu^{sw}\tau^{tv}) \\
&\quad \times (\frac{1}{6}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}).
\end{aligned}$$

We consider each of the terms in  $b_{\text{NP}}(\theta)$  separately under orthogonality:

$$\begin{aligned}
& - (\tau^{rs}\nu^{tu} + \nu^{rs}\tau^{tu})(\frac{1}{4}\lambda_{rstu} - \lambda_{rst/u} + \lambda_{rt/su}) \\
&= (\lambda_{11})^{-1}\lambda^{ab}(\frac{1}{2}\lambda_{11ab} - \lambda_{1ab/1} - \lambda_{11a/b}); \\
& - (\tau^{rs}\tau^{tu}\nu^{vw} + \tau^{rs}\nu^{tu}\tau^{vw} + \nu^{rs}\tau^{tu}\tau^{vw} - \tau^{rs}\nu^{tu}\nu^{vw} - \nu^{rs}\tau^{tu}\nu^{vw} - \nu^{rs}\nu^{tu}\tau^{vw}) \\
&\quad \times (\frac{1}{4}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}) \\
&= -(\lambda_{11})^{-2}\lambda^{ab}(\frac{1}{2}\lambda_{111}\lambda_{1ab} + \frac{1}{4}\lambda_{11a}\lambda_{11b} - \lambda_{1ab}\lambda_{11/1} + \lambda_{11/1}\lambda_{ab/1}) \\
&\quad - (\lambda_{11})^{-1}\lambda^{ab}\lambda^{cd}(\frac{1}{2}\lambda_{11a}\lambda_{bcd} + \frac{1}{4}\lambda_{1ab}\lambda_{1cd} - \lambda_{11a}\lambda_{bc/d} - \lambda_{1ab}\lambda_{1c/d} + \lambda_{11/a}\lambda_{bc/d}); \\
& - (\tau^{ru}\tau^{sw}\nu^{tv} + \tau^{ru}\nu^{sw}\tau^{tv} + \nu^{ru}\tau^{sw}\tau^{tv} - \tau^{ru}\nu^{sw}\nu^{tv} - \nu^{ru}\tau^{sw}\nu^{tv} - \nu^{ru}\nu^{sw}\tau^{tv}) \\
&\quad \times (\frac{1}{6}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \lambda_{rs/t}\lambda_{uv/w}) \\
&= -(\lambda_{11})^{-2}\lambda^{ab}(\frac{1}{2}\lambda_{11a}\lambda_{11b} - \lambda_{11a}\lambda_{11/b} - \lambda_{11a}\lambda_{1b/1}) \\
&\quad - (\lambda_{11})^{-1}\lambda^{ab}\lambda^{cd}(\frac{1}{2}\lambda_{1ac}\lambda_{1bd} - \lambda_{1ac}\lambda_{bd/1}).
\end{aligned}$$

The resulting formula for  $b_{\text{NP}}(\theta)$  in the presence of orthogonal nuisance parameters is

$$\begin{aligned}
b_{\text{NP}}(\theta) &= (\lambda_{11})^{-1}\lambda^{ab}(\frac{1}{2}\lambda_{11ab} - \lambda_{1ab/1} - \lambda_{11a/b}) \\
&\quad - (\lambda_{11})^{-2}\lambda^{ab}(\lambda_{111}\lambda_{1ab} + \frac{3}{4}\lambda_{11a}\lambda_{11b} \\
&\quad \quad - \lambda_{11a}\lambda_{11/b} - \lambda_{11a}\lambda_{1b/1} - \lambda_{1ab}\lambda_{11/1} + \lambda_{11/1}\lambda_{ab/1}) \\
&\quad - (\lambda_{11})^{-1}\lambda^{ab}\lambda^{cd}(\frac{1}{2}\lambda_{11a}\lambda_{bcd} + \frac{1}{4}\lambda_{1ab}\lambda_{1cd} + \frac{1}{2}\lambda_{1ac}\lambda_{1bd} \\
&\quad \quad - \lambda_{11a}\lambda_{bc/d} - \lambda_{1ab}\lambda_{1c/d} - \lambda_{1ac}\lambda_{bd/1} + \lambda_{11/a}\lambda_{bc/d}).
\end{aligned}$$

Just as for  $g_{\text{NP}}(\theta)$  in the case of orthogonal nuisance parameters, we see that  $b_{\text{NP}}(\theta)$  involves multiple sums over the indices for the nuisance parameters, so  $b_{\text{NP}}(\theta)$  can be expected to be large when the number of nuisance parameters is large.

An interesting feature emerges from comparing the formulas for  $g_{\text{NP}}(\theta)$  and  $b_{\text{NP}}(\theta)$  in the orthogonal nuisance parameter case. While the expression for  $g_{\text{NP}}(\theta)$  involves a double sum over the indices for the nuisance parameters, the expression for  $b_{\text{NP}}(\theta)$  involves both double and quadruple sums. Consequently, we might reasonably expect the ratio  $b_{\text{NP}}(\theta)/b_{\text{INF}}(\theta)$  to grow more rapidly with the number of nuisance parameters than does the ratio  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta)$ . This phenomenon is apparent in Example 1, for which  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta) = 3q/2$ . It turns out that  $b_{\text{INF}}(\theta) = n^{-1}\frac{1}{3}$  and  $b_{\text{NP}}(\theta) = n^{-1}(q^2 + q)$ , so  $b_{\text{NP}}(\theta)/b_{\text{INF}}(\theta) =$

$3(q^2 + q)$ . In this example, the ratio  $b_{\text{NP}}(\theta)/b_{\text{INF}}(\theta)$  grows quadratically with the number of nuisance parameters, while the ratio  $g_{\text{NP}}(\theta)/g_{\text{INF}}(\theta)$  only grows linearly.

## 6. DECOMPOSITIONS FOR OTHER PIVOTS

So far, our focus has been on inference based on an adjusted version of the signed root likelihood ratio statistic; however, other pivots that are asymptotically standard normal also find widespread use, notably the Wald-type pivots based on the difference  $\hat{\psi} - \psi$  and the score-type pivots based on the derivative  $M_1(\psi) = dM(\psi)/d\psi = L_1\{\tilde{\theta}(\psi)\}$ . DiCiccio et al. (2015) provide analysis of circumstances where inference, such as  $p$ -values, obtained by bootstrapping various first-order asymptotically equivalent pivots will agree to higher-order with that obtained from the signed root statistic. It is of interest to assess the impact that nuisance parameters have on higher-order adjustments obtained by Cornish-Fisher transformation to these other pivots. We examine the structure of these adjustments in terms of the quantities  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$ , to allow explicit comparisons with inference based on  $R(\psi)$ .

Let  $T(\psi)$  denote an asymptotically standard normal pivot, and let its cumulants be denoted by  $\kappa_1, \kappa_2$ , etc. Typically, the mean  $\kappa_1$  and skewness  $\kappa_3$  are of order  $O(n^{-1/2})$ , while the variance  $\kappa_2 = 1 + O(n^{-1})$ ; the fourth and higher-order cumulants are of order  $O(n^{-1})$  or smaller. Central to higher-order inference based on  $T(\psi)$  is the Cornish-Fisher transformation  $T - \frac{1}{6}\kappa_3 T^2 - \kappa_1 + \frac{1}{6}\kappa_3$ , which has the standard normal distribution to error of order  $O(n^{-1})$ . The Cornish-Fisher transformation of  $R(\psi)$  agrees with the  $R^*(\psi)$  statistic to error of order  $O(n^{-1})$ . The adjustment terms  $\frac{1}{6}\kappa_3$  and  $-\kappa_1 + \frac{1}{6}\kappa_3$  that appear in the Cornish-Fisher transformation depend on  $\theta$ , so they would need to be estimated to achieve higher-order inference in practice. An interpretation of the adjustment made by the Cornish-Fisher transformation is that whether or not a mean adjustment suffices to make the desired correction hinges on the order of  $\kappa_3$ . This is an important factor differentiating the signed root statistic from other asymptotically standard normal pivots.

We report  $\kappa_1$  and  $\kappa_3$  for some common choices of  $T(\psi)$ . For  $T(\psi) = R(\psi)$ , we have seen that  $\kappa_1 = -g_{\text{INF}}(\theta) - g_{\text{NP}}(\theta) + O(n^{-1})$ ; in this case,  $\kappa_3 = O(n^{-1})$ . Consequently, higher-order inference based on  $R(\psi)$  requires estimation of  $\kappa_1$  only, and estimation of  $\kappa_3$  is not necessary.

To report  $\kappa_1$  and  $\kappa_3$  for other pivots  $T(\psi)$ , it is convenient to introduce one further asymptotic quantity in addition to  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$ . This quantity is  $d \equiv d(\theta) = -\eta^{1/2} \frac{1}{6} \lambda^{1r} \tau^{st} \lambda_{rst}$ , which arises quite naturally from the profile log-likelihood function. It turns out that the third derivative of the profile log-likelihood function evaluated at  $\hat{\psi}$  is  $M_3(\hat{\psi}) = \eta^{3/2} 6d + O_p(n^{1/2})$ . The quantity  $d$  is also related to Efron's (1987) asymptotic adjustments  $a_0$  and  $c_q$ , which were discussed by DiCiccio & Efron (1996):  $d = 2a_0 + c_q$ . Furthermore, in terms of  $g_{\text{INF}}(\theta)$ ,  $g_{\text{NP}}(\theta)$ , and  $d$ , the mean of  $\hat{\psi}$  is  $E(\hat{\psi}) = \psi - (2g_{\text{INF}}(\theta) + g_{\text{NP}}(\theta) - d)\eta^{-1/2} + O(n^{-3/2})$ .

A key property of the quantity  $d$  is that it is the same whether there are no nuisance parameters or there are orthogonal nuisance parameters. In both cases, the formula for  $d$  becomes  $d = -(-\lambda_{11})^{-3/2} \frac{1}{6} \lambda_{111}$ . Thus,  $d$  is similar to  $g_{\text{INF}}(\theta)$ : we would not expect  $d$  to grow with the number of nuisance parameters. The quantity  $d$  does differ from  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  in one important respect: while  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  are invariant under reparameterizations  $\theta = (\psi, \phi) \rightarrow \{g(\psi), h(\psi, \phi)\}$ , where  $\phi = (\theta^2, \dots, \theta^d)$  contains the nuisance parameters and  $g(\psi)$  is a monotonically increasing function,  $d$  does not enjoy the property of invariance.



We next consider the Wald statistic with observed information,  $T(\psi) = (\hat{\psi} - \psi)/(-\hat{L}^{11})^{1/2}$ , and the Wald statistic with expected information,  $T(\psi) = (\hat{\psi} - \psi)/(-\hat{\lambda}^{11})^{1/2} = (\hat{\psi} - \psi)\hat{\eta}^{1/2}$ . The distributions of these pivots are the same to error of order  $O(n^{-1})$ . For both Wald statistics,  $\kappa_1 = -\{g_{\text{INF}}(\theta) + g_{\text{NP}}(\theta) + d\} + O(n^{-1})$  and  $\kappa_3 = -6d + O(n^{-1})$ . Consequently, the Wald statistics are similar to the signed root of the likelihood ratio statistic in that nuisance parameters affect the higher-order adjustment terms through  $g_{\text{NP}}(\theta)$ , which is involved in  $\kappa_1$ .

Finally, we consider the score statistic with observed information,  $T(\psi) = M_1(\psi)(-\hat{L}^{11})^{1/2}$ , and the score statistic with expected information,  $T(\psi) = M_1(\psi)(-\hat{\lambda}^{11})^{1/2} = M_1(\psi)\hat{\eta}^{1/2}$ . Just as for the Wald statistics discussed above, the distributions of these pivots agree to error of order  $O(n^{-1})$ ; for these score statistics,  $\kappa_1 = -\{g_{\text{INF}}(\theta) + g_{\text{NP}}(\theta) - 2d\} + O(n^{-1})$  and  $\kappa_3 = 12d + O(n^{-1})$ . Again, nuisance parameters influence the higher-order adjustment terms through  $g_{\text{NP}}(\theta)$ , which is a component of  $\kappa_1$ .

An important property of the profile log-likelihood function  $M(\psi)$  is that the expectation of the profile score is  $E\{M_1(\psi)\} = -\eta^{1/2}g_{\text{NP}}(\theta) + O(n^{-1})$ . Thus,  $E\{M_1(\psi)\}$  is of order  $O(1)$ ; the expectation of the profile score does even vanish asymptotically. Adjusted profile likelihood is discussed in the Appendix. Most of the adjustment functions  $B(\psi)$  that have been proposed to construct an adjusted profile log-likelihood  $\bar{M}(\psi) = M(\psi) + B(\psi)$  have the property that  $E\{B_1(\psi)\} = \eta^{1/2}g_{\text{NP}}(\theta) + O(n^{-1})$ , so the expectation of the adjusted profile score is  $E\{M_1(\psi)\} = O(n^{-1})$ , which does vanish asymptotically.

For  $T(\psi) = \bar{R}(\psi) = \text{sgn}(\bar{\psi} - \psi)[2\{\bar{M}(\bar{\psi}) - \bar{M}(\psi)\}]^{1/2}$ , as detailed in the Appendix, we have  $\kappa_1 = -g_{\text{INF}}(\theta) + O(n^{-1})$  and  $\kappa_3 = O(n^{-1})$ . Thus, at order  $O(n^{-1/2})$ , the difference between the distribution of  $\bar{R}(\psi)$  and the standard normal distribution depends on  $g_{\text{INF}}(\theta)$ , a term which is the same whether there are no nuisance parameters present or there are orthogonal nuisance parameters. Consequently, we expect the difference between the distribution of  $\bar{R}(\psi)$  and the standard normal distribution not to grow inordinately as the number of nuisance parameters increase.

Similar comments apply to Wald statistics and score statistics based on the adjusted profile log-likelihood function. For example, for  $T(\psi) = (\bar{\psi} - \psi)\{-\bar{M}_{11}(\bar{\psi})\}^{1/2}$ , we have  $\kappa_1 = -\{g_{\text{INF}}(\theta) + d\} + O(n^{-1})$  and  $\kappa_3 = -6d + O(n^{-1})$ , while for  $T(\psi) = M_1(\psi)\{-\bar{M}_{11}(\bar{\psi})\}^{-1/2}$ , we have  $\kappa_1 = -\{g_{\text{INF}}(\theta) - 2d\} + O(n^{-1})$  and  $\kappa_3 = 12d + O(n^{-1})$ .

Implementation of higher-order inference to error of order  $O(n^{-1})$  requires that we estimate the adjustment terms  $\frac{1}{6}\kappa_3$  and  $-\kappa_1 + \frac{1}{6}\kappa_3$ ; we might, for example, use plug-in estimates or derive estimates from a simulation procedure such as the parametric bootstrap. If these adjustment terms change rapidly with the value of the parameter  $\theta$ , then there is greater scope for error in the estimation process than if possible if the adjustment terms are stable across  $\theta$  values. This observation points to the use of asymptotically standard normal pivots  $T(\psi)$  that are derived from the adjusted profile log-likelihood function, since the adjustment terms for such pivots depend only on  $g_{\text{INF}}(\theta)$  and  $d$ . If the adjustment terms are small in magnitude, then they are unlikely to vary unduly with  $\theta$ , and the adjustments can be estimated more reliably. Situations can arise, as is the case in the normal regression example, that the quantity  $g_{\text{NP}}(\theta)$  is large yet it remains constant with respect to  $\theta$ . In these circumstances, the need to use the adjusted profile log-likelihood is not so pressing; indeed, for the normal regression model, the parametric bootstrap affords exact inferences, except for simulation error. Since such situations are not commonplace, there is strong motivation for using generally procedures that ensure the magnitudes of the adjustment terms are controlled. However, it could be useful to develop conditions that easily identify models, such as the

normal linear regression model, for which the adjustment terms, especially  $g_{\text{NP}}(\theta)$ , are constant or nearly so, since, in such models, the benefit of using adjusted profile likelihood for accurate inference is not so pronounced and procedures based on the regular profile likelihood are likely to suffice.

## 7. DISCUSSION

Accurate inference on a scalar interest parameter  $\psi$  in the presence of a nuisance parameter may be obtained using the signed root likelihood ratio statistic  $R(\psi)$ . A computationally intensive, but analytically simple, approach bases the inference on a bootstrap estimate of the sampling distribution of  $R(\psi)$ , constructed by fixing the nuisance parameter at its observed constrained maximum likelihood value. Alternatively, inference can be based on a standard normal approximation to the sampling distribution of an analytically adjusted version of  $R(\psi)$ . For this latter approach, the gold standard is represented by Barndorff-Nielsen's  $R^*$  statistic. The adjustment made by this statistic may be decomposed into a sum of two terms. These adjustments  $\text{INF}(\psi)$  and  $\text{NP}(\psi)$  are determined to second order,  $O_p(n^{-1})$ , by their expectations.

We have provided an explicit evaluation of these expectations, allowing new theoretical interpretation of the relative importance of the two adjustments and to the intrinsic difficulty of the inference problem within any specified model.

In particular, quantifying the dependence of the expectations on the nuisance parameter provides insight to circumstances where the bootstrap and analytic approaches might be expected to perform well in terms of accuracy, even in high dimensional problems and with small sample sizes. We have demonstrated that within a particular model, the importance of the nuisance parameter adjustment may depend not only on the structure of the model, as expressed by the nuisance parameter dimension, but the parameter values themselves. In key problems, dependence lies only on the parameter dimension. Calculation of the approximations  $g_{\text{INF}}(\theta)$  and  $g_{\text{NP}}(\theta)$  of  $E\{\text{INF}(\psi)\}$  and  $E\{\text{NP}(\psi)\}$  involves only evaluation of expectations of low order log-likelihood derivatives, and has been demonstrated to give useful theoretical insight to the degree of the adjustment to the signed root statistic  $R(\psi)$  given by the statistic  $R^*(\psi)$  for any specified inference problem, and therefore to the likely value in use of  $R^*(\psi)$  or bootstrapping as a means of improving accuracy.

We note that empirical estimation of the means, through the bootstrap principle of estimation of the nuisance parameter, furnishes a simple procedure for adjustment of the signed root likelihood ratio statistic. A thorough analysis of this empirical adjustment method for the purposes of inference with higher-order accuracy, as well as a comparison of such an empirical adjustment method with alternative approximations, is beyond the scope of this paper.

## APPENDIX

**Adjusted profile likelihood.** There have been many suggestions to replace the usual profile likelihood function  $M(\psi)$  by an adjusted version  $\bar{M}(\psi) = M(\psi) + B(\psi)$ , where  $B(\psi)$  is an adjustment function whose derivatives with respect to  $\psi$  are of order  $O_p(1)$ . The likelihood ratio statistic based on the adjusted profile likelihood is  $\bar{W}(\psi) = 2\{\bar{M}(\bar{\psi}) - \bar{M}(\psi)\}$ , where  $\bar{\psi}$  is the point at which  $\bar{M}(\psi)$  is maximized. The signed root of the likelihood ratio statistic based on the adjusted profile likelihood is  $\bar{R}(\psi) = \text{sgn}(\bar{\psi} - \psi)\{\bar{W}(\psi)\}^{1/2}$ .

Following our previous notation, we write  $B_1(\psi) = \partial B(\psi)/\partial\psi$ ,  $B_{11}(\psi) = \partial^2 B(\psi)/\partial\psi^2$ , etc. Let  $\beta_1 = E\{B_1(\psi)\}$ ,  $\beta_{11} = E(B_{11})$ , etc.; these quantities are assumed to be of order  $O(1)$ . Further, let  $b_1 = B_1(\psi) - \beta_1$ ,  $b_{11} = B_{11}(\psi) - \beta_{11}$ , etc., with these quantities assumed to be of order  $O_p(n^{-1/2})$ . Assume also that the joint cumulants of  $nb_1$ ,  $nb_{11}$ ,  $l_r$ ,  $l_{rs}$ , etc. are of order  $O(n)$ .

In many instances, the adjustment function  $B(\psi)$  has been proposed to take into account the effect of nuisance parameters for inference about  $\psi$ ; see, notably, Cox & Reid (1987), Barndorff-Nielsen (1983), Skovgaard (1996), Severini (1998), DiCiccio & Martin (1993), Barndorff-Nielsen & Chamberlin (1994). These adjustment functions have the effect of reducing the expectation of the profile score from order  $O(1)$  to order  $O(n^{-1})$ . Specifically, these functions have  $\beta_1 = \rho + O(n^{-1})$ , where  $\rho = -\eta\lambda^{1r}\nu^{st}(\frac{1}{2}\lambda_{rst} + \lambda_{rs,t})$ . Since, in general,  $E\{M_1(\psi)\} = -\rho + O(n^{-1})$ , it follows that  $E\{\bar{M}_1(\psi)\} = O(n^{-1})$ : see McCullagh & Tibshirani (1990), DiCiccio et al. (1996).

For a general adjustment function  $B(\psi)$ , DiCiccio & Stern (1994b) showed that  $\bar{R}(\psi) = \eta^{1/2}\{\bar{R}_1 + \bar{R}_2 + O_p(n^{-3/2})\}$ , where  $\bar{R}_1 = R_1 = -\lambda^{1r}l_r$  and  $\bar{R}_2 = R_2 - \lambda^{11}\beta_1$ ; in particular,  $\bar{R}(\psi) = R(\psi) + \eta^{-1/2}\beta_1 + O_p(n^{-1})$ . Below, we use this result with a particular adjustment function to obtain a representation of the nuisance parameter adjustment  $\text{NP}(\psi)$ , from which  $E\{\text{NP}(\psi)\}$  is then determined to  $O(n^{-1})$ . Combined with (3), this enables calculation to  $O(n^{-1})$  of  $E\{\text{INF}(\psi)\}$ .

**Expectations of Adjustments.** We have,

$$\begin{aligned} E\{\text{NP}(\psi)\} + E\{\text{INF}(\psi)\} &= -E\{R(\psi)\} + O(n^{-1}) \\ (4) \quad &= -\eta^{1/2}\{\lambda^{1r}\lambda^{st}\lambda_{rs,t} + \frac{1}{2}\lambda^{1r}\tau^{st}\lambda_{rs,t} + \frac{1}{2}\lambda^{1r}\lambda^{st}\lambda_{rst} + \frac{1}{3}\lambda^{1r}\tau^{st}\lambda_{rst}\} + O(n^{-1}). \end{aligned}$$

It is easily seen that  $\text{NP}(\psi)$  and  $\text{INF}(\psi)$  are of the form  $\text{NP}(\psi) = E\{\text{NP}(\psi)\} + O_p(n^{-1})$  and  $\text{INF}(\psi) = E\{\text{INF}(\psi)\} + O_p(n^{-1})$ . Here we develop explicit approximations for  $E\{\text{NP}(\psi)\}$  and  $E\{\text{INF}(\psi)\}$ .

The quantity  $\text{NP}(\psi)$  is related to the modified profile likelihood of Barndorff-Nielsen (1983), an adjusted profile likelihood which reduces the bias of the profile score. Following Sartori et al. (1999) and Pierce & Bellio (2006), we have that, up to an additive constant, the log modified profile likelihood is

$$\begin{aligned} L^{MP}(\psi) &= -R(\psi)\text{NP}(\psi) - \{R(\psi)\}^2/2 \\ &= -R(\psi)\text{NP}(\psi) - M(\hat{\psi}) + M(\psi) \\ &= -\frac{1}{2}\{R(\psi) + \text{NP}(\psi)\}^2 + O_p(n^{-1}). \end{aligned}$$

The modified profile likelihood therefore corresponds to an adjustment function of the form  $B(\psi) = -R(\psi)\text{NP}(\psi)$ . Further, the signed square root of the modified profile likelihood ratio statistic is equivalent, to  $O_p(n^{-1})$ , to  $R(\psi) + \text{NP}(\psi)$ , as noted by Sartori et al. (1999). The general result of DiCiccio & Stern (1994b) then gives  $\text{NP}(\psi) = \eta^{-1/2}\beta_1 + O_p(n^{-1})$ .

Observing that  $R(\psi) = (\hat{\psi} - \psi)\hat{\eta}^{1/2} + O_p(n^{-1/2})$  and  $\text{NP}(\psi) = \text{NP}(\hat{\psi}) + O_p(n^{-1})$ , we have

$$\begin{aligned} L^{MP}(\psi) &= -R(\psi)\text{NP}(\psi) - M(\hat{\psi}) + M(\psi) \\ &= (\psi - \hat{\psi})\hat{\eta}^{1/2}\text{NP}(\hat{\psi}) - M(\hat{\psi}) + M(\psi) + O_p(n^{-1}), \end{aligned}$$

and differentiation with respect to  $\psi$  yields

$$L_1^{MP}(\psi) = \hat{\eta}^{1/2}\text{NP}(\hat{\psi}) + M_1(\psi) + O_p(n^{-1/2}) = \eta^{1/2}\text{NP}(\psi) + M_1(\psi) + O_p(n^{-1/2}).$$

Since (see, for example, DiCiccio et al., 1996)  $E\{L_1^{MP}(\psi)\} = O(n^{-1})$  and  $E\{M_1(\psi)\} = -\rho + O(n^{-1/2})$ , it follows that

$$E\{\text{NP}(\psi)\} = \eta^{-1/2}\rho + O(n^{-1}) = -\eta^{1/2}\lambda^{1r}\nu^{st}(\lambda_{rs,t} + \frac{1}{2}\lambda_{rst}) + O(n^{-1}),$$

so that  $\beta_1 = \eta^{1/2}E\{\text{NP}(\psi)\} + O(n^{-1/2}) = -\eta\lambda^{1r}\nu^{st}(\lambda_{rs,t} + \frac{1}{2}\lambda_{rst}) + O(n^{-1/2})$ . It follows from (4) that

$$E\{\text{INF}(\psi)\} = \eta^{1/2}\lambda^{1r}\tau^{st}(\frac{1}{2}\lambda_{rs,t} + \frac{1}{6}\lambda_{rst}) + O(n^{-1}).$$

We observe also that this analysis confirms  $E\{\text{NP}(\psi)\} = \eta^{-1/2}\beta_1 + O(n^{-1}) = \text{NP}(\psi) + O_p(n^{-1})$ , as noted earlier.

## REFERENCES

- (1) BARNDORFF-NIELSEN, O. E. (1983) On a formula for the conditional distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–65.
- (2) BARNDORFF-NIELSEN, O. E. (1986) Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–22.
- (3) BARNDORFF-NIELSEN, O. E. & CHAMBERLIN, S. R. (1994) Stable and invariant adjusted directed likelihoods. *Biometrika* **81**, 485–99.
- (4) BARNDORFF-NIELSEN, O. E. & COX, D. R. (1994) *Inference and Asymptotics*. London: Chapman and Hall.
- (5) BRAZZALE, A. R. & DAVISON, A. C. (2008) Accurate parametric inference for small samples. *Stat. Sci.* **23**, 465–84.
- (6) BRAZZALE, A. R., DAVISON, A. C. & REID, N. (2007) *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge: Cambridge University Press.
- (7) COX, D. R. & REID, N. (1987) Parameter orthogonality and approximate conditional inference (with discussion). *J.R. Statist. Soc. B* **53**, 79–109.
- (8) DICICCIO, T. J. & EFRON, B. (1996) Bootstrap confidence intervals (with discussion). *Stat. Sci.* **11**, 189–228.
- (9) DICICCIO, T. J. & MARTIN, M. A. (1993) Simple modifications for signed roots of likelihood ratio statistics. *J.R. Statist. Soc. B* **55**, 305–16.
- (10) DICICCIO, T. J. & STERN, S. E. (1994a) Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J.R. Statist. Soc. B* **56**, 397–408.
- (11) DICICCIO, T. J. & STERN, S. E. (1994b) Constructing approximately standard normal pivots from signed roots of adjusted likelihood ratio statistics. *Scand. J. Statist.* **21**, 447–60.
- (12) DICICCIO, T. J. & YOUNG, G. A. (2008) Conditional properties of unconditional parametric bootstrap procedures for inference in exponential families. *Biometrika* **95**, 747–58.
- (13) DICICCIO, T. J., MARTIN, M. A. & STERN, S. E. (2001) Simple and accurate one-sided inference from signed roots of likelihood ratios. *Can. J. Statist.* **29**, 67–76.
- (14) DICICCIO, T. J., MARTIN, M. A., STERN, S. E. & YOUNG, G. A. (1996) Information bias and adjusted profile likelihood. *J.R. Statist. Soc. B* **58**, 189–203.
- (15) DICICCIO, T. J., KUFFNER, T. A., YOUNG, G. A. & ZARETZKI, R. (2015) Stability and uniqueness of p-values for likelihood-based inference. To appear in *Statistica Sinica*.

- (16) EFRON, B. (1987) Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.* **82**, 171–200.
- (17) FRASER, D. A. S. (1990) Tail probabilities from observed likelihoods. *Biometrika* **77**, 65–76.
- (18) LAWLEY, D. N. (1956) A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303.
- (19) LEE, S. M. S. & YOUNG, G. A. (2005) Parametric bootstrapping with nuisance parameters. *Stat. Prob. Letters* **71**, 143–53.
- (20) MCCULLAGH, P. & TIBSHIRANI, R. (1990) A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. B* **52**, 325–44.
- (21) PIERCE, D. A. & BELLIO, R. (2006) Effects of the reference set on frequentist inferences. *Biometrika* **93**, 425–38.
- (22) PIERCE, D. A. & PETERS, D. (1992) Practical use of higher-order asymptotics for multiparameter exponential families (with discussion). *J.R. Statist. Soc. B* **54**, 701–38.
- (23) SARTORI, N. (2003) Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika* **90**, 533–49.
- (24) SARTORI, N., BELLIO, R., SALVAN, A. & PACE, L. (1999) The directed modified profile likelihood in models with many nuisance parameters. *Biometrika* **86**, 735–42.
- (25) SEVERINI, T. A. (1998) An approximation to the modified profile likelihood function. *Biometrika* **85**, 403–11.
- (26) SEVERINI, T. A. (2000) *Likelihood methods in Statistics*. Oxford: Oxford University Press.
- (27) SKOVGAARD, I. M. (1996) An explicit large-deviation approximation to one-parameter tests. *Bernoulli* **2**, 145–65.
- (28) YOUNG, G. A. (2009) Routes to higher-order accuracy in parametric inference. *Aust. N.Z. J. Stat.* **51**, 115–26.

DEPARTMENT OF SOCIAL STATISTICS, CORNELL UNIVERSITY, ITHACA, NEW YORK 14853, U.S.A.

*E-mail address:* tjd9@cornell.edu

DEPARTMENT OF MATHEMATICS, WASHINGTON UNIVERSITY IN ST. LOUIS, ST. LOUIS, MISSOURI 63130, U.S.A.

*E-mail address:* kuffner@math.wustl.edu

DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON, LONDON SW7 2AZ, U.K.

*E-mail address:* alastair.young@imperial.ac.uk